Letter to the Editor

Sofie K. Van Houcke, Stefan Van Aelst, Katleen Van Uytfanghe and Linda M. Thienpont*

Harmonization of immunoassays to the all-procedure trimmed mean – proof of concept by use of data from the insulin standardization project

Keywords: equivalence of measurement results; principal component analysis; split-sample measurements.

***Corresponding author: Linda M. Thienpont,** Harelbekestraat 72, 9000 Gent, Belgium, Phone: +32 9 2648104, Fax: +32 9 2648198, e-mail: linda.thienpont@ugent.be
Sofie K. Van Houcke, Katleen Van Uytfanghe and Linda M. Thienpont: Laboratory for Analytical Chemistry, Faculty of Pharmaceutical Sciences, Ghent University, Gent, Belgium
Stefan Van Aelst: Department of Applied Mathematics and Computer Science, Ghent University, Gent, Belgium

Today, numerous laboratory procedures are available for measurement of clinically relevant biomarkers. Evidence-based practice guidelines are considered of valuable assistance in medical decision-making, however, recommending uniform cut-offs is only meaningful provided different measurement procedures give equivalent results within clinically relevant constraints. To achieve this, trueness-based standardization is the ideal scenario. For a documentation of the efforts already achieved in this regard, we refer to the International System of Units (SI)-traceable reference materials and measurement procedures in the database of the Joint Committee for Traceability in Laboratory Medicine (http://www.bipm.org/jctlm). These apply for well-defined chemical entities or internationally recognized reference method-defined measurands, such as enzymes. For the more complicated components, particularly those present in the biological system as mixture, it is unlikely that the goal of SI-traceability will be accomplished in the short-term [1]. This was exactly the incentive to reflect on the utility of establishing a technical infrastructure based on other means than reference measurement procedures to accomplish harmonization with equivalency of measurement results [2, 3]. From a scrutiny of literature we were surprised to find that, already decades ago, the potential of the combination of split-sample multiple method comparison studies with statistical analysis of the data by principal component analysis (PCA) had been described [4–6]. These exemplary studies showed that PCA can compare differences and similarities between measurement procedures, provide a so-called composite reference value for each of the measured samples, so that they subsequently can be used as a valid basis for calibration and accomplishment of equivalent measurement results. This approach eliminates calibration biases among different measurement procedures, but does not establish a link to the truth. To rehearse this alternative to harmonization of measurements, we did a proof of concept study. We applied PCA to existing data from a split-sample method comparison study between immunoassays and an isotope dilution-mass spectrometry (ID-MS)-based candidate SI-traceable reference measurement procedure for serum or plasma insulin [7, 8]. The data set was of particular interest because it had been used before to prove the feasibility of SI-based standardization of insulin measurements.

The data were measurement results for 36 single-donor sera obtained by 11 immunoassays from nine manufacturers and ID-MS. The sera covered an insulin concentration range from 18 to 779 pmol/L (3.1 to 130 mU/L, according to the conversion factor 0.1667 adopted from http://diabetes.diabetesjournals.org/site/misc/SIunits.pdf). The uncertainty of the serum concentrations measured by ID-MS was approximately 4%. We applied PCA to derive the all-procedure trimmed mean (APTM) from the immunoassay results. For a detailed account of PCA, we refer to reference [9]. The assumption at the basis of the used PCA model was that the 11 immunoassays are different procedures that measure a given quantity with an unknown underlying reference value in the samples, and that the measured quantity values include a random and

systematic error component. Note that the systematic error or bias may consist of a fixed, method-specific part which expresses to what extent a procedure consistently over/underestimates the reference value, and a random, sample-specific part which is due to non-specificity, i.e., response to other substances than the target. With x_{ij} the measurement of the j th procedure for the i th sample, the PCA model can be written as $x_{ij}=\alpha_j+\beta_j f_i+\varepsilon_{ij}$, where the values f_i are the reference values for the samples. Note that in the absence of a reference measurement procedure the parameters β_j and reference values f_i in the above model can only be estimated relatively, i.e., up to an unknown scalar. We used a robust PCA procedure [10]. Note that the standard choice in PCA algorithms is to return estimates of the reference values f_i that are centered around zero. However, for our purposes these relative estimates can easily be adjusted to better match the centers of the results. We compared the APTM from PCA with the ID-MS results in a scatter plot and a percentage residual plot using weighted least squares regression analysis. We also did correlation analysis between the measurement results by the 11 immunoassays and the ID-MS results and APTM, respectively. For data analysis, we used Microsoft Excel®, except for weighted least squares regression analysis, for which we utilized CB-Stat.

Figure 1 compares the insulin concentrations of the panel samples calculated as APTM from PCA (concentrations in pmol/L) with those measured by ID-MS in a scatter diagram (A) and a percentage residual plot (B). The weighted least squares regression and correlation data were APTM=1.089×ID-MS+4.31 pmol/L and r^2=0.9970, with 95% of the percentage residuals within ±6.7%. Correlation analysis of the measurement results by the 11 immunoassays with ID-MS and the APTM gave a mean Pearson correlation coefficient r^2 of 0.9916 (range 0.9639–0.9992) and 0.9929 (0.9710–0.9994), respectively.

The data set from the insulin method comparison and standardization with an SI-traceable reference measurement procedure served as proof of concept for the use of the APTM from PCA as alternative to calibration. It allowed in particular the comparison of the quality of both calibration approaches. The validity of the statistical approach can be inferred from the excellent correlation of the APTM with ID-MS (r^2=0.9970). Moreover, the observation that 95% of the percentage residuals are within limits that correspond fairly well with the expected distribution assuming that the uncertainty for PCA is the same as for ID-MS values (√2*4%=5.7%) reflects that the APTM compensates well for the among-assay differences in specificity. This is an important aspect of the proof of concept, in view of the fact that sufficient specificity is inherent to the reference measurement procedure approach. The success of calibration against the APTM relies on the fact that the concerned immunoassays are in a mature status [see the high values for r^2 (>0.99) in the correlation analysis between the immunoassay results and the APTM]. It should be noted that the regression data (APTM vs. ID-MS) show a positive bias of the APTM vs. the reference measurement procedure, reflecting that only the latter establishes SI-traceability, vs. the APTM that is a statistical composite of the calibration status of the immunoassays it is based on. Therefore, it is important to consider that the continuity of calibration against the APTM must be ensured by overlapping measurements of the first calibration panel with the follow-up ones [2]. The panels may serve as predicate panels for

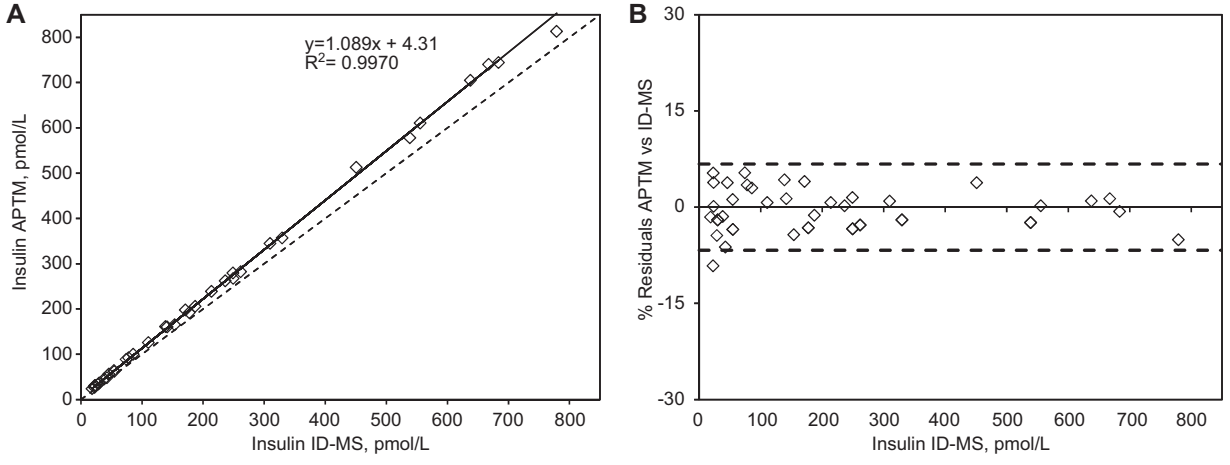

Figure 1 Comparison of insulin concentrations of the panel samples calculated as APTM from PCA (pmol/L), with those measured by ID-MS. (A) Scatter diagram (with dotted line: line of equality) and (B) percentage residual plot (with dotted lines: 1.96 * $SD_{\%\ residuals}$).

all measurement procedures that are entering the market after the first calibration.

In conclusion, we reiterated the great potential of the APTM derived by PCA to contribute to the harmonization of laboratory measurement procedures. Our study confirmed not only the validity of this statistical approach, but also showed that it results in an equivalent quality of calibration as the reference measurement procedure approach. Naturally, the more mature measurement procedures contribute to the APTM, the better the harmonization will be. Nevertheless, even when a first method comparison shows that harmonization is not yet feasible (gross differences between the results), it can be helpful for manufacturers towards the improvement of quality, so that harmonization may become feasible in a later stage.

Conflict of interest statement

Authors' conflict of interest disclosure: The authors stated that there are no conflicts of interest regarding the publication of this article.
Research funding: None declared.
Employment or leadership: None declared.
Honorarium: None declared.

Acknowledgments: The authors of the presented study are indebted to the members of the work group convened by the American Diabetes Association (see reference [7]) for kindly giving the permission to use the insulin data set. They also are grateful for the helpful advice given by D. Stöckl (STT-Consulting).

Received October 1, 2012; accepted October 12, 2012

References

1. Thienpont LM. Accuracy in clinical chemistry – who will kiss sleeping beauty awake? Clin Chem Lab Med 2008;46:1220–2.
2. Thienpont LM, Van Houcke SK. Traceability to a common standard for protein measurements by immunoassay for in-vitro diagnostic purposes. Clin Chim Acta 2010;411:2058–61.
3. Miller GW, Myers GL, Gantzer ML, Kahn SE, Schönbrunner ER, Thienpont LM, et al. Roadmap for harmonization of clinical laboratory measurement procedures. Clin Chem 2011;57:1108–17.
4. Lawton WH, Sylvestre EA, Young-Ferraro BJ. Statistical comparison of multiple analytic procedures: application to clinical chemistry. Technometrics 1979;21:397–409.
5. Carey RN, Wold S, Westgard JO. Principal component analysis: an alternative to "referee" methods in method comparison studies. Anal Chem 1975;47:1824–9.
6. Rymer JC, Sabatier R, Daver A, Bourleaud J, Assicot M, Bremond J, et al. A new approach for clinical biological assay comparison and standardization: application of principal component analysis to a multicenter study of twenty-one carcinoembryonic antigen immunoassay kits. Clin Chem 1999;45:869–81.
7. Miller GW, Thienpont LM, Van Uytfanghe K, Clark PM, Lindstedt P, Nilsson G, et al. Toward standardization of insulin immunoassays. Clin Chem 2009;55:1011–8.
8. Rodriguez-Cabaleiro D, Van Uytfanghe K, Stove V, Fiers T, Thienpont LM. Pilot study for the standardization of insulin immunoassays with isotope dilution-liquid chromatography/tandem mass spectrometry. Clin Chem 2007;53:1462–9.
9. Jolliffe IT. Principal component analysis, 2nd ed. New York: Springer-Verlag, 2002:487.
10. Hubert M, Rousseeuw PJ, Van Aelst S. High-breakdown robust multivariate methods. Stat Sci 2008;23:92–119.